Amendment to the Claims:

This listing of claims will replace all prior versions and listings of claims in the application:

Listing of Claims:

1. (currently amended) A computer-implemented method of determining a statistical model for predicting disease risk for a member of a population, comprising:

collecting, at at least one computing device, a plurality of sets of data, each of said sets of data associated with one member of said population, and comprising non-genetic, genetic data, and an indicator of disease status of said one member associated with said set;

storing, at at least one computing device, a candidate statistical model for calculating said disease risk as a function of non-genetic data, said candidate model dependent on a plurality of parameters;

~~determining, by at least one computing device, a plurality of weights, each one of said weights associated with one of said sets of data and indicating a statistical significance of said one of said sets of data, wherein weights associated with sets of said data having like genetic data are the same; and~~

optimizing, by at least one computing device, said parameters of said candidate model by fitting, wherein said fitting comprises:

calculating for each of said sets, a deviate of a predicted risk from an indicator of disease status for that set, said predicted risk predicted using said candidate model and non-genetic data in that set;

calculating a sum of weighted deviates for all of said sets, wherein each deviate is weighted in said sum by [[the]] a weight associated with, and indicating a statistical significance of, that set for which said each deviate has been calculated, and wherein the weights used to weight said deviates are determined with a constraint

that said weights associated with sets of said data having like genetic data are the same; and

minimizing said sum of weighted deviates to obtain optimized parameters, so that a risk calculated using said candidate model with said optimized parameters and non-genetic data associated with a particular member of said population is indicative of a disease risk to said particular member; and

providing said candidate model with said optimized parameters to a user to be used for calculating said disease risk.

2. (previously presented) The method of claim 1, wherein said deviate is a difference between said predicted risk and said indicator of disease status.

3. (previously presented) The method of claim 1, wherein each weighted deviate is obtained by multiplying the corresponding weight and a function of the corresponding deviate.

4. (previously presented) The method of claim 1, wherein said determining comprises:
grouping said collected data into groups such that all sets of data within each said group have like genetic data, one of said groups being a reference group which contains sets of data having genetic data like genetic data obtained from said member of said population; and
determining a group weight for each said group, whereby said group weight is the corresponding weight for each set of data within said each group.

5. (original) The method of claim 4, wherein the group weight of said reference group has a value of one and each of the other group weights has a value between zero and one.

6. (original) The method of claim 5, wherein said other group weights are optimized by minimizing a target function, said target function dependent on a plurality of residuals, one of said residuals for each of the data sets in said reference group.

7. (previously presented) The method of claim 6, wherein a residual for an *i*th one of said data sets in said reference group is the difference between a value of the indicator of disease status contained in said *i*th data set and the value of disease risk for the member associated with said *i*th data set, said value of disease risk calculated from said candidate model with said parameters optimized for a given set of group weights by fitting data sets in groups other than the reference group to said candidate model.

8. (original) The method of claim 7, wherein said target function is of the form:

$$f = \sum w_i \phi(r_i),$$

where

w_i is the corresponding weight for data set i; and
r_i is the residual for data set i.

9. (previously presented) The method of claim 1, wherein said non-genetic data comprises data indicative of time.

10. (original) The method of claim 9, wherein said candidate model is a Cox proportional hazard regression model.

11. (original) The method of claim 9, wherein said candidate model is a disease risk function of the form:

$$R(t) = 1 - \exp\{-\int_0^t h(u)du\},$$

where

$R(t)$ represents said disease risk at a given time t;

$h(u)$ is of the form:

$$h(u) = h_0(u)\exp(\sum_{1}^{n_c}\beta_i x_i)\,;$$

$h_0(u)$ is dependent only on u;

x_i is a variable indicative of a disease risk factor, said collected data containing a plurality of values of x_i;

β_i is a coefficient for x_i; and

n_c is the number of coefficients in said disease risk function.

12. (original) The method of claim 1, wherein said collecting comprises imputing missing data to said plurality of data sets.

13. (previously presented) The method of claim 1, wherein each one of said plurality of said weights is weighted by an adjustment factor.

14. (original) The method of claim 13, wherein an adjustment factor a_i for a data set obtained from a member i of said population is calculated as:

$$a_i = \frac{n_i^p}{n_i^s},$$

where n_i^p is the number of members in said population who share a same set of characteristics with said member i, and n_i^s is the number of members associated with said collected data who share said set of characteristics.

15. (original) The method of claim 14, wherein said set of characteristics comprises non-genetic factors.

16. (canceled)

17. (original) The method of claim 14, wherein said set of characteristics comprises both genetic and non-genetic factors.

18. (original) The method of claim 14, wherein said set of characteristics are selected from the group of age, gender, race, body mass index, smoking status, hypertension, cholesterol level, personal health history, and family health history.

19. (previously presented) The method of claim 1, comprising calculating said risk for said particular member of said population using said candidate model with said optimized parameters.

20. (previously presented) A computing system comprising at least one computing device, adapted for performing the method of any one of claims 1 to 19.

21. (currently amended) An article of manufacture comprising a computer readable medium embedded thereon computer executable instructions, which when executed by a computer causes said computer to determine a statistical model for predicting disease risk for a member of a population by

collecting a plurality of sets of data, each of said sets of data associated with one member of said population, and comprising non-genetic data, genetic data, and an indicator of disease status of said one member associated with said set;

storing a candidate statistical model for calculating said disease risk as a function of non-genetic data, said candidate model dependent on a plurality of parameters;

~~determining a plurality of weights, each one of said weights associated with one of said sets of data and indicating a statistical significance of said one of said sets of data, wherein weights associated with sets of said data having like genetic data are the same;~~

optimizing said parameters of said candidate model by fitting, wherein said fitting comprises:

calculating for each of said sets, a deviate of a predicted risk from an indicator of disease status for that set, said predicted risk predicted using said candidate model and non-genetic data in that set;

calculating a sum of weighted deviates for all of said sets, wherein each deviate is weighted in said sum by [[the]] a weight associated with, and indicating a statistical significance of, that set for which said each deviate has been calculated, and wherein the weights used to weight said deviates are determined with a constraint that said weights associated with sets of said data having like genetic data are the same; and

minimizing said sum of weighted deviates to obtain optimized parameters, so that a risk calculated using said candidate model with said optimized parameters and non-genetic data associated with a particular member of said population is indicative of a disease risk to said particular member; and

providing said candidate model with said optimized parameters to a user to be used for calculating said disease risk.

22. (previously presented) The method of claim 1, wherein each of said sets of data is indicative of a plurality of factors, and said collecting comprises:

determining a correlation between said plurality of factors;

grouping said factors into batches such that all factors in each said batch are correlated; and

imputing missing data for factors in one said batch at a time.

23. (previously presented) The method of claim 4, wherein said grouping comprises:

dividing said plurality of sets of data into two or more groups depending on data indicative of a non-genetic factor in each of said data sets;

determining if a criterion is met after said dividing, said criterion is evaluated based on genetic data in each of said data sets; and

when said criterion is not met, regrouping said plurality of sets of data back into one group.

24. (original) The method of claim 23, wherein said dividing is performed recursively on each group of a division.

25. (original) The method of claim 24, wherein divisions at different levels are made dependent on data indicative of different factors.

26. (original) The method of claim 25, wherein a branch of said recursive division is terminated at the level at which said criterion is not met.

27. (currently amended) A computer-implemented method of weighing a plurality of data sets, each one of said data sets associated with a member of a population, comprising:

weighing, by at least one computing device, each set of said plurality of data sets by a weight indicative of a representativeness of the member associated with said each set in said population, wherein a weight a_i for a data set obtained from a member i of said population is calculated as:

$$a_i = \frac{n_i^p}{n_i^s},$$

where n_i^p is the number of members in said population who share a same set of characteristics with said member *i*, and n_i^s is the number of members associated with said collected data who share said set of characteristics; ~~and~~

storing, at at least one computing device, said weight a_i in association with said data set obtained from said member *i*; and

providing said data sets and their associated weights to a user to be used for adjusting a statistical weight of each one of said data sets.

28. (currently amended) A computer-implemented method of determining a statistical model for predicting disease risk for a member of a population,

storing, at at least one computer, a plurality of statistical models, each for calculating said disease risk;

for each of said models, assessing, by at least one computer, a goodness of fit of data derived from a plurality of members of said population, said assessing comprising calculating

a deviate from an indicator of a disease status of each member by a predicted risk for that member, predicted using that model and non-genetic data associated with that member, and

a sum of weighted deviates, each deviate weighted by a weight reflecting genetic data associated with that member for whom that deviate is calculated; [[and]]

selecting the model that produces the lowest sum of weighted deviates as a risk prediction model for predicting said disease risk; and

providing said risk prediction model to a user to be used for predicting said disease risk.

29. (previously presented) The method of claim 28, wherein each of said models is dependent on a plurality of parameters, different ones of said models having different numbers of parameters.

30. (previously presented) The method of claim 28, wherein each of said models is dependent on a plurality of parameters, different ones of said models having an equal number of parameters.